Semantic Image Segmentation Using Oriented Pattern Analysis

Hui Ting Ong and Kai-Kuang Ma
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
htong@ntu.edu.sg, ekkma@ntu.edu.sg

***Abstract*—Semantic image segmentation is often served as an instrumental front-end pre-processing of many image processing and computer vision applications. Different from the existing semantic segmentation methods which mostly rely on database training, a simple and effective approach is proposed in this paper by employing the *oriented pattern* as a class-discriminative feature, besides the use of color and texture. The developed algorithm is to discriminate three specific semantic classes—*sky*, *foliage*, and *building*, which are commonly encountered in typical color images containing outdoor scene. In our approach, two maps will be generated from the input image: 1) the over-segmented map by using JSEG, and 2) the *oriented pattern map* by measuring the *orientation coherence* at each pixel position. Based on these maps, the *region coherence* of each over-segmented patch can be obtained by taking the average of all the orientation coherence values within the patch. The semantic class labeling for the patch can be simply conducted by checking the computed region coherence value against two empirically determined thresholds. Extensive simulation experiments have been conducted by testing various color images acquired from various sources, including the Microsoft Research Cambridge (MSRC) Object Recognition Image Database that contains manually-segmented regions as the ground truth. All experimental results clearly show that the proposed method is able to consistently deliver highly attractive performance.**

***Index Terms*—Semantic image segmentation; oriented pattern analysis; semantic class; object segmentation**

I. Introduction

Image segmentation is a challenging image processing task which aims to partition an image into multiple regions. Accurate and meaningful segmentation plays an important role as a pre-processing step in many image processing and computer vision applications. Conventional segmentation algorithms like N-cut [1], mean-shift [2] and JSEG [3] that perform segmentation based on color and texture features tend to yield over-segmentation results; consequently, a single object could be unnecessarily partitioned into several homogeneous regions. In order to obtain useful segmentation results that are meaningful to human visual perception, most of the existing semantic image segmentation methods rely on database training [4]-[8]. These approaches exploit a fixed model with the parameters tuned for each object category of interest based on a set of training images that contain the concerned object. However, the performance of these approaches intimately depends on the model selected and the database used.

Fig. 1. (Top Row) Three typical outdoor images that contain three frequently encountered semantic classes—sky, foliage, and building; (Bottom Row) their resulted oriented pattern maps through the measurement of orientation coherence, respectively.

The main objective of our work is to conduct semantic image segmentation focusing on three semantic classes that are commonly encountered in a typical image containing outdoor scene—that is, *sky*, *foliage*, and *building*. The proposed method exploits the *oriented pattern* [9] that is expected to be inherited in each semantic class as an effective class-discriminative feature, to discriminate the above-mentioned three semantic classes. For that, two maps are independently generated from the input image under semantic segmentation. One is the oriented pattern map which is generated by computing the orientation coherence at each pixel position with consideration of its neighboring gradients. The other is the JSEG-based over-segmented map, which is based on the homogeneity of color and texture of the input image. The generated two maps are then combined to compute the region coherence for yielding the final semantic image segmentation results.

The rest of the paper is organized as follows. Section II discusses the proposed semantic image segmentation approach in detail. Section III presents extensive simulation results and discussions. Finally, conclusions are drawn in Section IV.

II. Proposed Oriented-pattern-based Approach

A. Motivation

In conventional segmentation approaches (e.g., [3]), salient features like color and texture are commonly exploited for image segmentation. However, using these features alone might not be able to handle *semantic* object segmentation. For

978-1-4577-0031-6/11/$26.00 ©2011 IEEE

an illustration, refer to Fig. 1 (top row) and observe that both color and texture drastically change from region to region even within the same building; thus, it would be difficult to segment this building accurately as one object simply based on these two features. In this paper, it has been demonstrated that the *oriented pattern* [9] can be exploited as a *class-discriminative* feature for discriminating three specific semantic classes—*sky*, *foliage*, and *building* that are commonly encountered in those images containing outdoor scene. This is due to the intuition that each of the above-mentioned class should posses an inherited oriented pattern; for example, sky is normally appeared as a *homogeneous* region, foliage tends to yield an *incoherent* orientation pattern, while a building inclines to have more *coherent* orientation pattern.

To extract the class-discriminative feature for the image under semantic segmentation, an *oriented pattern map*, which has values ranging from 0 to 1 computed at each pixel position, is to be generated. In fact, the oriented pattern map used in our work was inspired from the orientation coherence map used in a fingerprint enhancement application [10]. Some examples of the oriented pattern maps are shown in Fig. 1 (bottom row). The first step of generating the oriented pattern map is to determine whether each pixel belongs to a homogeneous region. For homogeneous regions like sky, the pixel gradients should be fairly close to zero. For non-homogeneous regions covering the other two classes (i.e., *foliage* and *building*), the degree of orientation coherence [9] needs to be computed in order to discriminate these two classes.

B. Mathematical Foundation

Given an image $I(x,y)$, the gradient $G(x,y)$ at pixel position (x,y) can be computed along the x-axis direction and y-axis direction, respectively; i.e., $G_x(x,y) = \partial I(x,y)/\partial x$, $G_y(x,y) = \partial I(x,y)/\partial y$. As the gradient $G(x,y)$ is a vector, it can be represented as $G(x,y) = G_x(x,y) + jG_y(x,y)$, where $j = \sqrt{-1}$ is the complex imaginary. For ease of presentation, the function argument (x,y) of $G_x(x,y)$, $G_y(x,y)$ and other similar mathematical functions to be introduced will be dropped hereafter. To generate the oriented pattern map, the *orientation coherence* at each pixel position (x,y) will be computed by considering the gradient vectors within a local region (denoted by window W). However, it has been observed that two gradient vectors with completely opposite directions, in fact, indicate the same orientation direction. Hence, instead of using the gradient G for computing orientation coherence, a new quantity defined as the *squared gradient* (denoted by G_s) will be used; that is,

$$\begin{aligned} G_s &= (G_x + jG_y)^2 = \left(G_x^2 - G_y^2\right) + j\left(2G_xG_y\right) \\ &= G_{sx} + jG_{sy}. \end{aligned} \tag{1}$$

Or, in polar form,

$$G_s = \left(|G|e^{j\angle G}\right)^2 = |G|^2 e^{j2\angle G} = |G_s|e^{j\angle G_s}. \tag{2}$$

From (2), it can be immediately arrived at

$$\begin{cases} |G_s| = |G|^2 = G_x^2 + G_y^2. \\ \angle G_s = 2\angle G. \end{cases} \tag{3}$$

In other words, the magnitude of the squared gradient $|G_s|$ is simply the square of the magnitude of the gradient $|G|$; and the phase of the squared gradient G_s is twice that of the gradient G. Indeed, squaring the gradients has the effect of rotating two gradient vectors with opposite direction (i.e., $180°$ apart) and aligning them into the same direction. As a result, the permissible range for the phase of the squared gradient G_s is limited to $[0°, 180°]$.

A measurement of the *orientation coherence* of the squared gradient G_s at each pixel position (x,y) within the window W can be computed as [9]

$$Coh = \frac{\sqrt{(\sum_W \gamma_i G_{sx})^2 + (\sum_W \gamma_i G_{sy})^2}}{\sum_W \gamma_i \sqrt{G_{sx}^2 + G_{sy}^2}}, \tag{4}$$

where γ_i denotes the weighting factor being multiplied at pixel i for the weighted summation performed using a Gaussian function with standard deviation $\sigma = 5$. (Note that $\sigma = 5$ is empirically determined from experiments). The values of Coh, which range from 0 to 1, reflect how consistent the squared gradients within the window W are pointing in the same direction; the larger the value, the stronger the coherence. That is, if all the squared gradient vectors are pointing in the same direction, the modulus of the weighted sum of the vectors (i.e., the denominator) equals to the sum of the weighted modulus of the vectors (i.e., the numerator); hence, $Coh = 1$. On the other hand, if the squared gradient vectors are equally pointing in all directions, vectors will be canceled from each other; hence, $Coh = 0$. However, simply using the computed orientation coherence is not able to discriminate between the class *sky* and the class *building* since both classes yield high orientation coherence values. Hence, we need to further impose the homogeneity property on the generation of the oriented pattern map as follows: For those pixel positions with gradient magnitude close to zero (say, $|G| \leq 0.001$, as empirically determined in our work), their Coh values will be set to zero in the oriented pattern map without involving the computation of Coh using (4).

C. Proposed Semantic Segmentation Framework

The framework of the proposed semantic segmentation method is shown in Fig. 2, where an over-segmentation operation using JSEG [3] is performed to group pixels into regions or patches with homogeneous color and texture, followed by labeling each patch with a semantic label. Consider the pixels from each patch are most likely belong to the same semantic class, the patches (rather than the pixels) are now viewed as the elementary units; hence, it is meaningful to compute a new quantity, called the *region coherence*, as follows. For each patch P_j consisting of N pixels, the average of the orientation coherence values contained in the patch P_j will be computed; that is,

$$\overline{M}_j = \frac{1}{N} \sum_{(x,y)\in P_j} M(x,y), \tag{5}$$

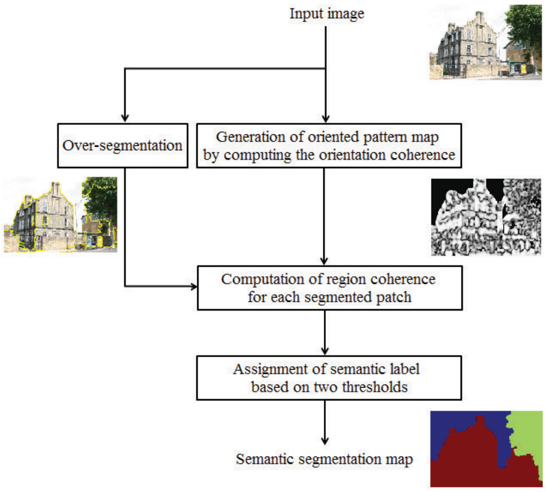

Fig. 2. The framework of the proposed semantic segmentation method.

where $M(x, y)$ denotes the orientation coherence value at pixel (x, y). Now, each patch is further classified into one of the three semantic classes based on the following criteria: If $\overline{M}_j < 0.2$, the patch is labeled as *sky*; if $0.2 \leq \overline{M}_j < 0.4$, as *foliage*; or, if $\overline{M}_j \geq 0.4$, as *building*. It should be noted that these two threshold values are empirically determined from our simulation experiments for the concerned three semantic classes.

III. Simulation Results

Extensive simulation experiments have been conducted to evaluate the performance of our proposed oriented-pattern-based semantic image segmentation method using a set of test images chosen from various sources. The Microsoft Research Cambridge (MSRC) Object Recognition Image Database, version 1.0 [11] contains 240 manually segmented and labeled photographs, covering 9 object classes (i.e., *building*, *grass*, *tree*, *cow*, *sky*, *aeroplane*, *face*, *car*, and *bicycle*). From the database, only the images that contain the three semantic classes (i.e., *sky*, *foliage* and *building*) are selected to test our algorithm. Some experimental results presented in Fig. 3 demonstrate how accurate the obtained semantic segmentation results compared with the ground truth provided in the MSRC database. Other images downloaded from Wikimedia and Google as well as some images acquired by using our own camera are also experimented with some results as documented in Fig. 4. One can see that our method consistently delivers fairly accurate segmentation results.

Overall, our approach works fairly well for most of the outdoor images containing *sky*, *foliage*, and *building*. However, it may fail in some special cases as expected; e.g., if a building has a large region containing almost no texture (say, a large concrete wall without windows), this part of the building could be mistakenly classified as *sky*. Another concern is that the size of window W used in the orientation coherence computation is required to be adjusted in case the concerned class object is small (i.e., scale issue); e.g., for a small building, a smaller window size (say, $\sigma = 3$) should be used.

IV. Conclusion

In this paper, a novel semantic image segmentation method by incorporating the oriented pattern as an effective class-discriminative feature is proposed to partition an image into generic classes. To demonstrate this core idea and develop our approach, three semantic classes are selected in this paper—namely, *sky*, *foliage* and *building*, which are commonly encountered in a typical image containing outdoor scene.

The key success of our approach lies in the use of the oriented pattern as another salient feature, besides color and texture as the two key features used in JSEG. The oriented pattern map is generated by computing the *orientation coherence* at each pixel position. By integrating the oriented pattern map with the JSEG over-segmentation map, the final semantic segmentation results can be obtained. To be more specific, the *region coherence* for each segmented patch from the JSEG map will be computed with reference to the corresponding orientation coherence values from the oriented pattern map simply by taking the average value of all orientation coherence values within the patch. The obtained region coherence is then checked against two empirically determined thresholds for the final classification or class labeling. The proposed method can achieve fairly accurate semantic segmentation results for the above-mentioned three semantic classes as verified through extensive simulation experiments and their obtained segmentation results.

References

[1] J. Shi and J. Malik, *Normalised Cuts and Image Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, 2000.

[2] D. Comaniciu and P. Meer, *Mean Shift: A Robust Approach Toward Feature Space Analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, 2002.

[3] Y. Deng and B. S. Manjunath, *Unsupervised Segmentation of Color-texture Regions in Images and Video*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 8, 2001.

[4] L. Yang, P. Meer and D. J. Foran, *Multiple Class Segmentation Using A Unified Framework over Mean-Shift Patches*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, Minnesota, USA, 2007.

[5] J. J. Lim, P. Arbelaez, C. Gu, and J. Malik, *Context by Region Ancestry*, International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009.

[6] Z. Tu, X. Chen, A. L. Yuille, S. C. Zhu, *Image Parsing: Unifying Segmentation, Detection, and Recognition*, International Conference on Computer Vision (ICCV), Nice, France, 2003.

[7] M. P. Kumar, P. H. S. Torr and A. Zisserman, *Obj Cut*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, California, USA, 2005.

[8] A. C. Berg, F. Grabler and J. Malik, *Parsing Images of Architectural Scenes*, International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, 2007.

[9] M. Kass and A. Witkin, *Analyzing Oriented Patterns*, Computer Vision, Graphics, and Image Processing, Vol. 37, No. 3, 1987.

[10] http://www.csse.uwa.edu.au/∼pk/Research/MatlabFns/

[11] http://research.microsoft.com/en-us/projects/ObjectClassRecognition/

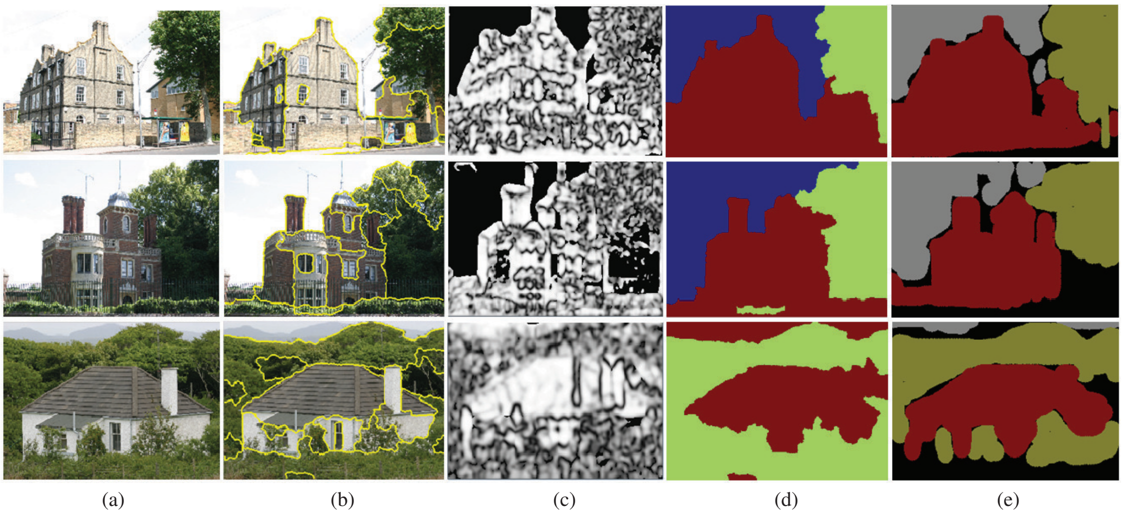

Fig. 3. Semantic image segmentation results by using our proposed method conducted on the test images obtained from the MSRC database (columns from left to right): (a) original images; (b) over-segmented images by using JSEG [3]; (c) computed oriented pattern maps; (d) segmented semantic classes—buliding (in red), foliage (in green), and sky (in blue); and (e) the ground truth segmentation results provided by the MSRC database, where the black regions denote void areas.

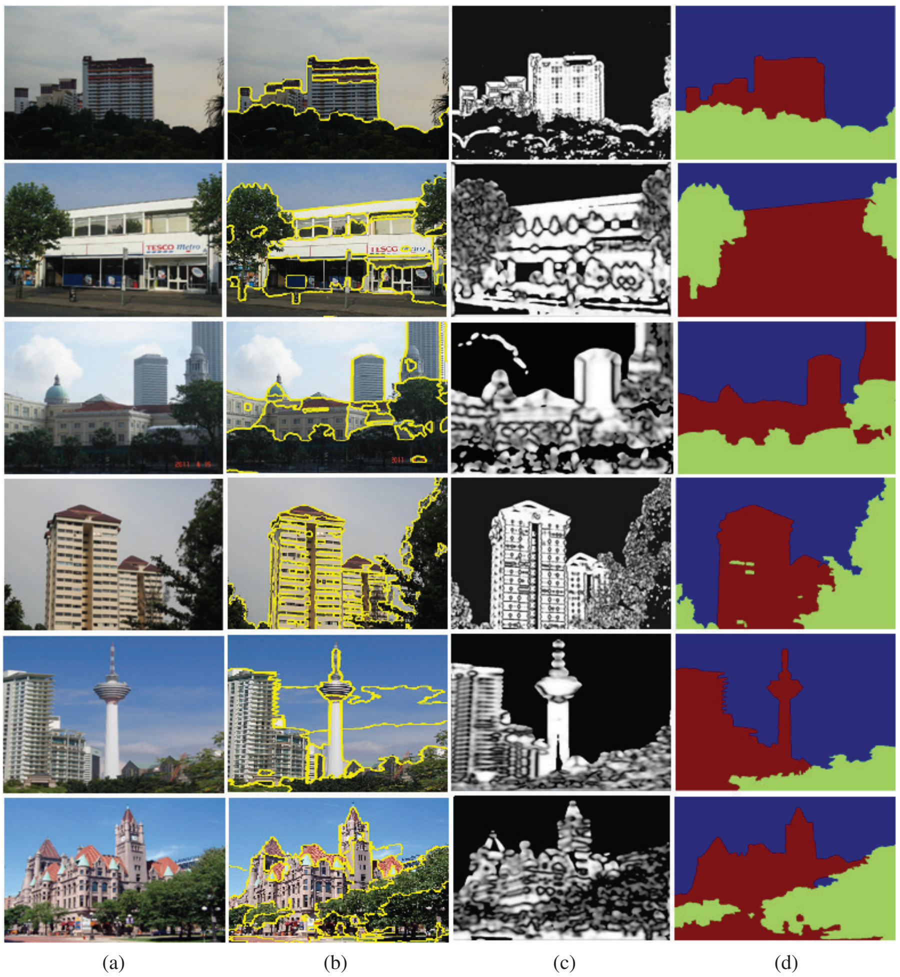

Fig. 4. Semantic image segmentation results by using our proposed method conducted on some test images obtained from Wikimedia, Google, and our own camera (columns from left to right): (a) original images; (b) over-segmented images by using JSEG [3]; (c) computed oriented pattern maps; and (d) segmented semantic classes—buliding (in red), foliage (in green), and sky (in blue).